

VIDEO ENCODING METHOD AND DEVICE

FIELD OF THE INVENTION

The present invention relates to a video encoding method provided for encoding an input image sequence consisting of successive groups of frames themselves subdivided into

5 blocks, said method comprising the steps of :

- preprocessing said sequence on the basis of a so-called content-change strength (CCS) computed for each frame by applying some predetermined rules ;
- estimating a motion vector for each block of the current frame ;
- generating a predicted frame using said motion vectors respectively associated to the
- 10 blocks of the current frame ;
- applying to a difference signal between the current frame and the last predicted frame a transformation sub-step producing a plurality of coefficients and followed by a quantization sub-step of said coefficients ;
- coding said quantized coefficients.

15 Said invention is for instance applicable to video encoding devices that require reference frames for reducing e.g. temporal redundancy (like motion estimation and compensation devices). Such an operation is part of current video coding standards and is expected to be similarly part of future coding standards also. Video encoding techniques are used for instance in devices like digital video cameras, mobile phones or digital video

20 recording devices. Furthermore, applications for coding or transcoding video can be enhanced using the technique according to the invention.

BACKGROUND OF THE INVENTION

In video compression, low bit rates for the transmission of a coded video sequence may be obtained by (among others) a reduction of the temporal redundancy

25 between successive pictures. Such a reduction is based on motion estimation (ME) and motion compensation (MC) techniques. Performing ME and MC for the current frame of the video sequence however requires reference frames (also called anchor frames). Taking MPEG-2 as an example, different frames types, namely I-, P- and B-frames, have been defined, for which ME and MC are performed differently : I-frames (or intra frames) are

30 independently coded, by themselves, without any reference to past or future frames (i.e. without any ME and MC), while P-frames (or forward predicted pictures) are encoded each one relatively to a past frame (i.e. with motion compensation from a previous reference

frame) and B-frames (or bidirectionally predicted frames) are encoded relatively to two reference frames (a past frame and a future frame). The I- and P-frames serve as reference frames.

In order to obtain good frame predictions, these reference frames need to be of high quality, i.e. many bits have to be spent to code them, whereas non-reference frames can be of lower quality (for this reason, a higher number of non-reference frames, B-frames in the case of MPEG-2, generally lead to lower bit rates). In order to indicate which input frame is processed as an I-frame, a P-frame or a B-frame, a structure based on groups of pictures (GOPs) is defined in MPEG-2. More precisely, a GOP uses two parameters N and M, where N is the temporal distance between two I-frames and M is the temporal distance between reference frames. For example, an (N,M)-GOP with N=12 and M=4 is commonly used, defining an " I B B B P B B B P B B B " structure.

Succeeding frames generally have a higher temporal correlation than frames having a larger temporal distance between them. Therefore shorter temporal distances between the reference and the currently predicted frame on the one hand lead to higher prediction quality, but on the other hand imply that less non-reference frames can be used. Both a higher prediction quality and a higher number of non-reference frames generally result in lower bit rates, but they work against each other since the frame prediction quality results from shorter temporal distances only.

However, said quality also depends on the usefulness of the reference frames to actually serve as references. For example, it is obvious that with a reference frame located just before a scene change, the prediction of a frame located just after the scene change is not possible with respect to said reference frame, although they may have a frame distance of only 1. On the other hand, in scenes with a steady or almost steady content (like video conferencing or news), even a frame distance of more than 100 can still result in high quality prediction.

From the above-mentioned examples, it appears that a fixed GOP structure like the commonly used (12, 4)-GOP may be inefficient for coding a video sequence, because reference frames are introduced too frequently, in case of a steady content, or at a unsuitable position, if they are located just before a scene change. Scene-change detection is a known technique that can be exploited to introduce an I-frame at a position where a good prediction of the frame (if no I-frame is located at this place) is not possible due to a scene change. However, sequences do not profit from such techniques if the frame content is almost completely different after some frames having high motion, with however no scene change at

all (for instance, in a sequence where a tennis player is continuously followed within a single scene). A previous European patent application, already filed by the applicant on October 14, 2003, with the filing number 03300155.3 (PHFR030124) has then described a new method for finding better reference frames. This method will be recalled below.

5 SUMMARY OF THE INVENTION

It is therefore the object of the invention to propose a video encoding method based on said previous method for finding good frames that can serve as reference frames, but allowing to reduce more noticeably the coding cost.

To this end, the invention relates to a video encoding method such as defined in
10 the introductory paragraph of the description and in which said CCS is used in said quantization sub-step for modifying the quantization factor used in said quantization sub-step, said CCS and said quantization factor increasing or decreasing simultaneously.

The invention also relates to a device for implementing said method.

BRIEF DESCRIPTION OF THE DRAWINGS

15 The present invention will now be described, by way of example, with reference to the accompanying drawings in which :

- Fig. 1 illustrates the rules used for defining, according to the description given in the previous European patent application cited above, the place of the reference frames of the video sequence to be coded ;

20 - Fig.2 shows an encoder carrying out the encoding method described in said previous European patent application, taking the MPEG-2 case as an example ;

- Fig.3 shows an encoder carrying out the encoding method according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

25 The document cited above describes a method for finding which frames in the input sequence can serve as reference frames, in order to reduce the coding cost. The principle of this method is to measure the strength of content change on the basis of some simple rules, such as listed below and illustrated in Fig.1, where the horizontal axis corresponds to the number of the concerned frame and the vertical axis to the level of the strength of content
30 change : the measured strength of content change is quantized to levels (for instance five levels, said number being however not a limitation), and I-frames are inserted at the

beginning of a sequence of frames having content-change strength (CCS) of level 0, while P-frames are inserted before a level increase of CCS occurs or after a level decrease of CCS occurs. The measure may be for instance a simple block classification that detects horizontal and vertical edges, or other types of measures based on luminance, motion vectors, etc.

5 An implementation of this previous method in the MPEG encoding case is described in Fig.2. The encoder comprises a coding branch 101 and a prediction branch 102. The signals to be coded, received by the branch 101, are transformed into coefficients and quantized in a DCT and quantization module 11, the quantized coefficients being then coded in a coding module 13, together with motion vectors MV. The prediction branch 102,
10 receiving as input signals the signals available at the output of the DCT and quantization module 11, comprises in series an inverse quantization and inverse DCT module 21, an adder 23, a frame memory 24, a motion compensation (MC) circuit 25 and a subtracter 26. The MC circuit 25 also receives the motion vectors MV generated by a motion estimation (ME) circuit 27 (many types of motion estimators may be used) from the input reordered frames (defined
15 as explained below) and the output of the frame memory 24, and these motion vectors are also sent towards the coding module 13, the output of which ("MPEG output") is stored or transmitted in the form of a multiplexed bitstream.

 The video input of the encoder (successive frames X_n) is preprocessed in a preprocessing branch 103. First a GOP structure defining circuit 31 is provided for defining
20 from the successive frames the structure of the GOPs. Frame memories 32a, 32b, are then provided for reordering the sequence of I, P, B frames available at the output of the circuit 31 (the reference frames must be coded and transmitted before the non-reference frames depending on said reference frames). These reordered frames are sent on the positive input of the subtracter 26 (the negative input of which receives, as described above, the
25 output predicted frames available at the output of the MC circuit 25, these output predicted frames being also sent back to a second input of the adder 23). The output of the subtracter 26 delivers frame differences that are the signals to be coded processed by the coding branch 101. For the definition of the GOP structure, a CCS computation circuit 33 is provided.

 It has then been observed that the higher the CCS – which can result from motion –
30 the less the viewer can really follow the presented video. It is consequently proposed, according to the present invention, to increase or decrease the quantization factor used in the module 11 as a function of the CCS - said CCS and the quantization factor increasing or decreasing simultaneously – which can be obtained by sending the output information of the CCS computation circuit towards the DCT and quantization module 11 of the coding branch.

As described in the conventional part of Fig.3 (said Fig.3 is introduced in the next paragraph in relation with the description of the invention), it is known, indeed, that the coding module 13 is in fact composed of a variable-length coding (VLC) circuit arranged in series with a buffer memory, the output of said memory being sent back towards a rate control circuit 133 for modifying the quantization factor.

According to the invention, and as shown in Fig.3 in which similar circuits are designated by the same references as in Fig.2, an additional connection 200 intended to allow to implement the proposed modification of quantization factor is provided between the CCS computation circuit 33 and the rate control circuit 133 and also between said circuit 33 and the DCT and quantization module 11 of the coding branch. This connection 200 extends two coding modes of the coding system, namely a so-called open-loop coding mode (without bit-rate control) and a closed loop coding mode (with bit-rate control).

In the open-loop coding mode for example, the quantizer settings are usually fixed. The resulting bit rate of the encoded stream is automatically lower for simple scenes (less residue needs to be coded) than for complex scenes (higher residue needs to be coded). Coding cases as described above, where the sequence contains high motion, result in complex scenes that are coded with high bit-rates. The bit-rates for the high-motion scenes can be reduced by higher quantization, thereby removing spatial details of these scenes that the observer cannot follow due to the motion. The quantization can be controlled by defining a quantization factor, q_{ccs} , which is a function of CCS and the original fixed quantizer factor, called q_{fixed} :

$$q_{ccs} = q_{fixed} + f(CCS),$$

where $f()$ is a function resulting in positive integers $0 \dots (q_{max} - q_{fixed})$ to increase q_{ccs} from q_{fixed} upto an allowed maximum q_{max} . Examples for $f()$ are $f1(CCS) = \text{round} (CCS * (q_{max} - q_{fixed}) / (CCS_{max}))$ or $f2(CCS) = \text{round} ((q_{max} - q_{fixed} + 1) ^ (CCS / CCS_{max}) - 1)$ for $CCS = 0$ to CCS_{max} .

In closed-loop coding, the quantization factor, q_{adapt} , is adapted in order to achieve a desired predefined bit rate. Bit-rate controllers that are required for closed-loop coding work basically with bit budgets and chose q_{adapt} based on the available budget. This means that the quantization factor q_{ccs} as described for open-loop coding can be used, and only q_{fixed} has to be replaced with q_{adapt} . Then, compared to an unmodified rate controller, the bit budget will increase with higher CCS, and these additional bits are automatically spent on frames with lower CCS, because the q_{adapt} value will decrease due to the increased bit budget.